Medical Informatics

MARK FRISSE, MD
Associate Editor

An Internet Primer: Resources and Responsibilities

MARK E. FRISSE, MD, ELIZABETH A. KELLY, MLS, EDWARD S. METCALFE

This is the third in a series of essays about networking and information services in academic medical centers. The first essay developed a philosophy of institutional data management and argued the necessity of integrated academic and clinical information systems. The second essay described the often conflicting needs of universities, health science schools, and hospitals, and presented one framework for organizing computer and network services across these settings. This essay has a more immediate and personal aim—to introduce the ever-expanding resources available through the Internet.

Many faculty and staff have access to networks for clinical or administrative reasons but do not understand that these networks (if properly designed and supported) can provide them with new ways of communicating with colleagues, remaining current on topics of professional interest, and obtaining programs and information from a wide range of international resources.[1] This essay is not intended for the experienced user, nor can it substitute for the many comprehensive Internet guides available in print or electronic form.[2-4] Instead, it is intended to help the uninitiated develop a framework for relating to a new means of academic communication—a starting point for lifelong involvement in an increasingly necessary and enjoyable technology.

For the sake of simplicity, actual commands and naming conventions are applicable to UNIX-based systems and are denoted in italics. Although the syntax may differ, the examples can easily be extended for users of other types of computer systems. This essay contains no illustrations; readers are urged to use the commands described here to see for themselves how the software systems discussed will look on their screens.

IDENTIFYING YOURSELF AND FINDING OTHERS

Those new to network-based computing are often discouraged by the simple matter of identification, conventions for which are still evolving. The general Internet address scheme is easily understood and most parties can be reached by the address syntax: *name@host.institution.domain.* Actual addresses, however, are difficult to guess. The six domains, such as education (*edu*) and private companies (*com*), are well defined, but insti-

tution names are shortened in a variety of ways, and individual host computers may be designated in any way or ways their administrators choose. A still greater problem arises in the "name" portion. Some systems at our institution identify clients in straightforward ways: for example,

mail frisse@medicine.WUSTL.EDU or
mail Mark.Frisse@medicine.WUSTL.EDU

will allow one to correspond with the first author of this paper. However, many administrative colleagues at the same institution have more cryptic addresses (e.g., C8413GA@WUMVS.BITNET). For effective communication, it seems imperative to identify oneself in a clear, straightforward way and to include a complete electronic mail address on any written correspondence that may lead to an exchange of electronic mail.

It is equally important to be able to find the addresses of others. There are a number of programs one can access to help to identify hostnames; most (e.g., nslookup, whois,netfind) are complex. If one has a good guess of the hostname and is only searching for the correct user name, the "finger" command is quite useful. One can, for example, find the home address of the first author by the command:

finger frisse@medicine.WUSTL.EDU or
finger Mark.Frisse@medicine.WUSTL.EDU

(case is not important). Some facilities have remarkably useful and "intelligent" ways of dealing with ambiguity. The command:

finger frosse@medicine.WUSTL.EDU

draws the response "????????." However, when sent to Stanford's CAMIS facility (where the author also has an account),

finger frosse@CAMIS.stanford.edu

returns 14 possible CAMIS users; the first item on the list is "frisse" (the command:

mail frisse@CAMIS.stanford.edu

automatically forwards items to the author's computer in St. Louis). A sound institutional information policy should include both clear protocols for selecting user names and useful software for ensuring that people with electronic mail accounts can be easily reached by their colleagues.

ANONYMITY: FTP AND LISTSERVS

There are times when one must remain anonymous. For instance, to obtain Internet statistics or an Internet reading list from Washington University's WUARCHIVE resource, one places an anonymous request through the FTP (File Transfer Protocol) package. FTP enables one to move through the directory structure of another computer connected to the Internet, and to transfer files—both programs and documents—to and from it as permitted. Many public FTP sites such as WUARCHIVE offer their resources to those who log in with the name "anonymous" and enter their full Internet address as a password (e.g, metcalfe@medicine.WUSTL.EDU). In this way, the command: *ftp ftp.sura.net* starts an FTP session on a remote computer called "ftp.sura.net," and allows access to such useful documents as the "Internet/Bitnet Health Sciences Resources" list compiled by Lee Hancock of the Dykes Library at the University of Kansas Medical Center.[5]

There are other times when one wishes one had been anonymous. Because a single Internet address can represent a group of individuals rather than one person, it is easy to broadcast personal notes in an embarrassing way. The academic health science librarians' distribution mailings, for example, give the impression that a simple *reply* command will send a message only to the individual sender; in actuality, the reply is sent to every health science librarian on the discussion list. As a more general instance, many facilities have distribution lists called "LISTSERVS." This software makes a great distinction between the address used to subscribe to a distribution and the address used to post a message to every member on the distribution list. To subscribe to the "NSF Grants and Contracts Forum," one issues the command:

mail LISTSERV@JHUVM.BITNET

and enters in the body of the message: *subscribe GRANTS-L.* The same message posted to the bulletin board itself would not subscribe, but would instead notify every reader on the list of one's ignorance about the confusing LISTSERV protocols.[6]

To avoid such embarrassment, social protocols for electronic mail should adhere to those honored for written correspondence.[7-8] Members of the authors' organization are urged never to forward personal mail without the consent of the writer; politeness is no less important in the ready domain of online editing and network postage.

Specific technical aspects of electronic mail, document exchange, and file transfer will evolve rapidly as true multimedia mail systems become more commonplace, as better automatic methods to exchange mail between different systems develop, and as programs to perform these tasks become easier to use. But the basic issues of identification, access, and social responsibility will remain relatively constant.

RESOURCE DISCOVERY

In a 1991 contribution to this column, Daniel Masys discussed the potential of the Internet to provide a wide range of resources relevant to health care professionals. At the time his article was written, most of these resources had to be accessed through FTP, electronic mail distribution lists, or cumbersome bulletin board programs. These tools can serve well for discussion and file transfer among known addresses; however, they offer little for surveying or discovering new resources.

To allow for use of the Internet's broader resources, electronic mail and FTP have been enhanced. For instance, LISTSERV software can automate a complicated response to a structured electronic mail message. The NSF uses LISTSERV software to provide automatic updates on grant programs rather than a forum for discussion. Similarly, the National Center for Biotechnology Information offers "BLAST" software that transmits GenBank queries and reports through electronic mail.

A few navigational aids have been developed among the more than 1,000 archives accessible by anonymous FTP. Generally, an "Index," "README," or "Is-IR" document may be downloaded and read as an introduction to an archive's contents. Also, broadly similar directory naming throughout the Internet enables one to browse more efficiently. These local guides and facilities do not, however, solve all accessibility problems. One effort to make the range of Internet resources more available to users is "archie." The Archie server, based at McGill University in Montreal and mirrored at servers around the world, maintains an index of every directory and file name available on known anonymous FTP archives. Archie supports string-based searches of its index and provides the complete Internet address and directory path for each matching item.

As a measure of scale and demand, updates and inquiries to McGill's archie server recently constituted 50% of all Montreal-bound Internet traffic.[9] Such traffic has strained the physical computing resources of the McGill host and of those that duplicate its service. To lessen the increase in response time and refused queries, a more efficient interface was sought. Initially, one could only access the archie servers directly, through the telnet application, which enables interactive sessions with a remote host. Telnet clients have long been used by research libraries and commercial databases to

support bibliographic searches over a network. A great deal of important data is presented for flexible querying in this way.[10] However, maintaining a constant connection and opening files and command interpreters for each session quickly exhausts the host computer. In response, the providers of archie turned to the Prospero system, which, through leaner communications, more efficient storage of directory names, and other means, has eased the burden on archie servers. In addition, clients who use Prospero[11] to access archie can integrate the transfer of files through FTP.

Notably, Prospero was developed for the scale and use of the contemporary Internet. It is an important, if inconspicuous, part of a suite of programs enabling new forms and breadth of information dissemination. These programs do not replace existing facilities such as electronic mail, telnet, and FTP, nor are they incompatible with existing information stores; rather, they integrate and augment these facilities in a redoubled effort to fulfill the educational and collaborative promise of a worldwide network.

THE INTERNET GOPHER

Gopher is the most popular new program in the public domain for storage and retrieval of information accessible over the Internet.[12] It was developed in 1991 at the University of Minnesota to offer seamless access to information produced and maintained by disparate campus departments. Individual departments could set up and maintain Gopher servers on whatever computers they might have. All of this information was presented in the form of directories on a top-level Gopher server accessible to individuals from most types of computers throughout the campus. Gopher has spread beyond the University of Minnesota to innumerable sites worldwide, each of which maintains local directories and pointers to elsewhere in "gopherspace."

All of gopherspace is navigated in the same way. One begins a session at one or another "home" Gopher server. Each server is equally accessible, and points more or less directly to all others. The command: *gopher medicine.WUSTL.EDU* initiates a session at the Washington University medical library's Gopher server is equally accessible, and points more or less directly to all others. The command:

gopher medicine.WUSTL.EDU

initiates a session at the Washington University medical library's Gopher server. Information on this Gopher archive, as on all others, is presented as a list of directories (marked with a "/") containing files or further directories. Directories often point to other Gopher sites. One may move through these directories and across the Internet by selecting directories and retracing selections. Any directory view can be annotated with a bookmark, which stores the path taken and enables direct access thereafter. Through the automated use of FTP, files may be viewed within the application and saved. Gopher also offers indexes (marked with a "<?>"). These may be "whois" phone and mail lists, or, more often, full-text document indexes supporting boolean and truncated queries. Indeed, most directory and file names in the worldwide Gopher system may be searched through a single index, "Veronica," the Gopher equivalent to archie. Veronica is generally referenced through a directory labeled "Other Gopher and Information Servers" (see, for instance, the server at *gopher.tc.umn.edu*).

The Gopher server software has proven useful both for small laboratories disseminating focused information and for large computer resources wishing to integrate campus-wide and nationwide information. Smaller groups tend to emphasize material and directory structures tied closely to their own operations, while larger resources take on the formidable task of providing access to many information resources and maintaining the currentness and relevance of these resources. All too often, these groups don't realize that while it is technically simple to establish a Gopher server, it takes effort to keep the server useful and current. Maintaining a list of conferences, schedules, and addresses on a Gopher server requires the coordination of a wide range of administrative and intellectual roles. As trained providers of a variety of information needs, librarians can be useful in coordinating these activities.

Organization of a Gopher server is equally crucial. The options presented on the screen in the Gopher servers of academic medical centers tend to be based either on source or on topic. Source-based Gopher servers simplify curatorial responsibilities, as each source can be maintained by a different party. (Indeed, some sources may be other Gopher servers anywhere on the Internet.) What this organizational form gains in ease of maintenance can be offset by user confusion. Where, for example, can one find a list of conferences related to biomedical engineering? In the "School of Medicine" menu item, the "School of Engineering" item, or in the "Institute for Biomedical Computing" item? Well-supported Gopher servers can minimize this confusion by including "pointer" menu items in every relevant category or by effective indexing, but both are imperfect solutions at best. Topic-based Gopher servers—if properly conceived—are easier to use but more difficult to maintain. If a Gopher server has an accomplished librarian–editor and a high degree of administrative support, a topic-based Gopher can be effective. Still other Gopher servers represent mixtures of source-based and topic-based Gophers. The top level generally offers a menu of general topics (e.g., "About this Gopher Server") and sources (e.g., "School of Medicine," "Hospital Administration," "University of Minnesota Gopher"). Within each source-based menu item, one is presented with a topic-based menu structure that is both clear and feasible to maintain.

WIDE-AREA INFORMATION SERVERS AND WORLDWIDE WEB

Gopher's many gateways and facilities provide an overarching directory structure from which other resource discovery programs depart. The protocol that supports full-text indexes in Gopher is actually a subset of another popular program, Wide Area Information Servers (WAIS). As the name implies, WAIS has a wide-area provenance distinct from Gopher's campus beginnings. Thinking Machines, Apple Computer, Dow Jones & Co., and KPMG Peat Marwick collaborated on the program to offer a single interface for querying and viewing results from any variety of geographically distributed free and commercial databases. WAIS clients[13] allow users to construct a natural language query and pose it to any number of selected sources simultaneously. Each question is responded to with a relevance-weighted list of appropriate documents from all source databases. These documents (which include multiple media) can be opened and saved directly.

This process rewards careful composition of questions and database sets. Stored questions may then be run periodically and modified accordingly. However, WAIS source databases have grown too numerous, and their names too cryptic, to support direct database selection. A "directory of servers" (quake.think.com), with a brief description of each, is now offered as a separate database. Also, a topical hierarchical menu of WAIS sources is under construction as a Gopher site.

World-wide Web (W3) takes a somewhat different approach. One begins within a document (again, this may be

multimedia). Phrases within the document are underlined, indicating links to further contextually appropriate information, which may be stored anywhere on the Internet and accessed with a single click. Linked information may be another W3 document (even a specific location within another W3 document), complete with further links; alternatively, links may open any of the network tools discussed above. It is intended that users toggle between indexes and documents—between requests to W3 servers and to those of Gopher, WAIS, archie, Hytelnet, and other software systems.

Each new link in the globally accessible W3 is authorized in negotiation with the publisher of an existing W3 document. This discrete development, which originated in the collaborative aims of the European Particle Physics Lab, has resulted in a conspicuously uneven resource. However, the usefulness of W3 does not rely on the completeness of its native information. A graphical client based on W3 software—such as Mosaic or, to a lesser extent, Cello[14]—excels in the forms of presentation, annotation, and navigation it permits. One may compose an individualized starting document of any complexity, framing links to specific Gopher sites, WAIS questions, and more. For now, W3 lacks a simple document editor; documents and links must be translated into the W3's Hypertext Markup Language (HTML). As clients and editors evolve, W3, which may be tailored to accommodate the unique needs of users, may increasingly serve the information needs generated by well-organized, distributed research collaborations.

Gopher, WAIS, and World-Wide Web continue to grow. More information and gateways are added each day. Likewise, the underlying software is developing in response to meet criticisms. As Gopher clients and servers convert to Gopher+ software, more complex client–server communication will allow more file formats and alternative directory views.[15] At the Center for Networked Information Development Research, WAIS is moving toward complete compliance with Z39.50 protocols, the emerging standard in library and commercial databases. World-Wide Web is developing a more finely grained information space, with varying access privileges and facilities for communicating with Internet archives.[16]

SUMMARY

The ease with which individuals can access the Internet and with which institutions can make information available on the Internet explains the exponential growth of this national resource. Once one accomplishes the difficult task of installing network services and establishing an ongoing mechanism for their support, it is relatively simple to use software systems such as those described in this article to gainfully traverse the Internet for a wide range of professional activities. But, as we have discussed, every step of the process, from simple naming conventions to organizations and ongoing maintenance of network-based information services, should proceed only after careful consideration of a network growing hourly in complexity. Despite the power of the technology available on one's desktop, one can often be frustrated by the small decisions: what is my colleague's email address? How can I most effectively find relevant information on home health care software? How should I organize a gopher server? When is WAIS preferable to Gopher or W3? Who will help me learn more?

The process comes full circle back to academic medical institutions. The usefulness of the Internet hinges upon the policies these institutions create to aid the organization and dissemination of medical information, and in the means they use to make their constituents aware of the pitfalls and potentials of various technologies. Expertise in the provision and support of networks must be complemented by expertise in the maintenance and dissemination of biomedical knowledge bases. The central intellectual issue from before the introduction of network-based resource, remains unchanged: how does one properly classify, archive, and disseminate knowledge so that it may best serve the needs of others? Librarians, administrators, scholars, clinicians, and students must work together, all contributing their best to the design and dissemination of Internet resources.

Dr. Frisse is associate dean for academic information management and director, School of Medicine Library and Biomedical Communications Center; Ms. Kelly is associate librarian, and Mr. Metcalfe is systems support technician, School of Medicine Library and Biomedical Communications Center; all at Washington University School of Medicine, St. Louis, Missouri.

Notes and References

1. Parties interested in public access to the Internet can avail themselves of a growing number of commercial services. Examples Church, Virginia (800) 488-6383; Sprint's Sprintlink in Herndon, Virginia (703) 904-2230; and Netcom of San Jose, California (408) 554-8649. Recent compilations can be found in a wide range of trade journals and in the popular press (e.g., "Basic Gear Gives Access to Network of Networks," *New York Times*, August 31, 1993, C3).
2. Krol, E. *The Whole Internet User's Guide and Catalog.* Sebastapol, California: O'Reilly and Associates, 1992.
3. Electronic Frontier Foundation. *Big Dummy's Guide to the Internet,* 1993. Available by anonymous FTP from ftp.eff.org in the directory /pub/EFF/papers.
4. Current bibliographies are posted on the Internet; many are available by anonymous FTP from infolib.murdoch.edu.au in the directory /pub/bib/.
5. This file is in the directory /pub/nic/ under the name "medical.resources." The editor can be reached at Le07144 @Ukanvm.cc.ukans.edu)
6. Among the valuable LISTSERVS are GRANTS-L at LISTSERV@JHUVM. BITNET (NSF Grants and Contracts), NIH-GUIDE at LISTSERV@TCSVM. BITNET (NIH Guide List), NIHDOC-L at LISTSERV@LSUVM.BITNET (NIH Guide List), NIHDIS-L at LISTSERV @JHUVM.BITNET (discussion), and NIHGUIDE at LISTSERV@UMAB.BITNET (NIH listing of available grants and contracts).
7. Rinaldi, A. H. (Rinaldi@ACC.FAU:EDU). *The Net: User Guidelines and Netiquette.* Available by anonymous FTP on many Internet servers, including nic.sura.net in the directory /pub/nic/internet.literature/ under the name "netiquette.txt."
8. Shapiro, N., et al. *Toward an Ethics and Etiquette for Electronic Mail.* Santa Monica, California: Rand Corporation, 1985. Available by anonymous FTP from rand.org in the directory pub/old/reports under the name "r-3283.1pr."
9. Emtage, A., and Deutsch, P. archie—An Electronic Directory Service for the Internet. *Proceedings of the 1992 USENIX Conference,* pp. 93–110. Berkeley, CA: USENIX Association, 1992.
10. A list of libraries and databases accessible through telnet is provided by the Hytelnet service, which indexes the Internet address and often idiosyncratic commands for each of these resources. Hytelnet browsers for most platforms are available by anonymous FTP from ftp.usask.ca in the directory /pub/Hytelnet/. One may also access this service through the telnet program by entering the command: *telnet access.usask.ca.*
11. A Prospero archie client is offered in XWindows and command line versions. These are available by anonymous FTP from many archives, including archie.ans.net in the directory /pub/. The archie service is also accessible through gateways provided in Gopher, WAIS, and World-Wide Web (described below).

12. Gopher clients (as well as servers) for most platforms are available by anonymous FTP from boombox.micro.umn.edu in the directory /pub/gopher/.
13. WAIS clients for Macintosh, Windows, etc., are available by anonymous FTP from ftp.think.com in the directory /wais/. WAIS also runs on many hosts; to check, enter *swais*, at the command line. One may also log on to a server run by Thinking Machines, Inc. with the command: *telnet quake.think.com.* Enter *wais* at the *login:* prompt, and continue as directed. The first set of menus offers WAIS as an option.
14. Mosaic for XWindows and Macintosh is available by anonymous FTP from ftp.ncsa.uiuc.edu in the directory Mac/Mosaic/. Cello, an integrated W3, Gopher, FTP, Telnet, CSO, and UseNet browser for Windows is available by anonymous FTP from fatty.law.cornell.edu in the directory pub/LII/ under the name "cello.zip"
15. Anklesari, F., et al. Gopher+: Upward Compatible Enhancements to the Internet Gopher Protocol (July 1993). Available by anonymous FTP from boombox.micro.umn.edu in the directory /pub/gopher/gopher_protocol/gopher+ under the name "gopher+.txt."
16. Berners-Lee, T. The World-Wide Web Initiative (1993) Available by anonymous FTP from info.cern.ch in the directory /pub/www/doc/.